THE SCIENCE OF SELF-ORGANIZATION AND ADAPTIVITY

Francis Heylighen,

Center "Leo Apostel", Free University of Brussels, Belgium

Summary: The theory of self-organization and adaptivity has grown out of a variety of disciplines, including thermodynamics, cybernetics and computer modelling. The present article reviews its most important concepts and principles. It starts with an intuitive overview, illustrated by the examples of magnetization and Bénard convection, and concludes with the basics of mathematical modelling. Self-organization can be defined as the spontaneous creation of a globally coherent pattern out of local interactions. Because of its distributed character, this organization tends to be robust, resisting perturbations. The dynamics of a self-organizing system is typically non-linear, because of circular or feedback relations between the components. Positive feedback leads to an explosive growth, which ends when all components have been absorbed into the new configuration, leaving the system in a stable, negative feedback state. Non-linear systems have in general several stable states, and this number tends to increase (bifurcate) as an increasing input of energy pushes the system farther from its thermodynamic equilibrium. To adapt to a changing environment, the system needs a variety of stable states that is large enough to react to all perturbations but not so large as to make its evolution uncontrollably chaotic. The most adequate states are selected according to their fitness, either directly by the environment, or by subsystems that have adapted to the environment at an earlier stage. Formally, the basic mechanism underlying self-organization is the (often noise-driven) variation which explores different regions in the system's state space until it enters an attractor. This precludes further variation outside the attractor, and thus restricts the freedom of the system's components to behave independently. This is equivalent to the increase of coherence, or decrease of statistical entropy, that defines self-organization.

1. Introduction

Science, and physics in particular, has developed out of the Newtonian paradigm of mechanics. In this world view, every phenomenon we observe can be reduced to a collection of atoms or particles, whose movement is governed by the deterministic laws of nature. Everything that exists now has already existed in some different arrangement in the past, and will continue to exist so in the future. In such a philosophy, there seems to be no place for novelty or creativity.

Twentieth century science has slowly come to the conclusion that such a philosophy will never allow us to explain or model the complex world that surrounds us. Around the middle of the century, researchers from different backgrounds and disciplines started to study phenomena that seemed to be governed by inherent creativity, by the spontaneous appearance of novel structures or the autonomous adaptation to a changing environment. The different observations they made, and the concepts, methods and principles they developed, have slowly started to coalesce into a new approach, a science of self-organization and adaptation. The present article will first present a quick, intuitive overview of these

developments, and then delve deeper into the abstract concepts and principles that came out of them.

2. The science of self-organization: a historical sketch

2.1. The thermodynamic paradox

The spontaneous emergence of new structures is easy to observe, both in the laboratory and in our day-to-day world. Perhaps the most common example is crystallization, the appearance of a beautifully symmetric pattern of dense matter in a solution of randomly moving molecules. A different example is the Bénard phenomenon, the appearance of a pattern of hexagonal cells or parallel rolls in a liquid heated from below. More complicated examples are certain chemical reactions, such as the Belouzov-Zhabotinsky reaction or the "Brusselator", where it suffices to constantly pump a number of ingredients into a solution in order to see the development of dazzling spirals of pulsating color.

What these examples have in common is *self-organization*: the appearance of structure or pattern without an external agent imposing it. It is as if the system of molecules arranges itself into a more ordered pattern. It was already noted that such a phenomenon contradicts the mechanistic world view. But it does not fit into our intuitive picture of the world either. If a system, such as a flower, a building or a watch, shows organization we tend to assume that someone or something must have arranged the components in that particular order. If we cannot find a person responsible for the design, we are tempted to attribute it to some unknown, intelligent force. This intuition is confirmed by the second law of thermodynamics, which says that in a system left to itself entropy (disorder) can only increase, not diminish. So, the first step to explain self-organization must be to reconcile it with thermodynamics.

In the example of crystallization, the solution is simple. The randomly moving molecules which become fixed within the crystalline structure pass on the energy of their movement to the liquid in which they were dissolved. Thus, the decrease in entropy of the crystal is compensated by an increase in the entropy of the liquid. The entropy of the whole, liquid and crystal together, effectively increases.

In the case where the self-organizing system does not reach an equilibrium, the solution is less obvious. The Belgian thermodynamicist Ilya Prigogine received a Nobel Prize for his investigation, starting in the 1950s, of that problem. Together with his colleagues of the "Brussels School" of thermodynamics, he has been studying what he called *dissipative structures*. These are patterns such as the Bénard cells or the Brusselator, which exhibit dynamic self-organization. Such structures are necessarily open systems: energy and/or matter are flowing through them. The system is continuously generating entropy, but this entropy is actively *dissipated*, or exported, out of the system. Thus, it manages to increase its own organization at the expense of the order in the environment. The system circumvents the second law of thermodynamics simply by getting rid of excess entropy. The most obvious examples of such dissipative systems are living organisms. Plants and animals take in energy and matter in a low entropy form as light or food. They export it back in a high entropy form, as waste products. This allows them to reduce their internal entropy, thus counteracting the degradation implied by the second law.

2.2. Principles of self-organization

The export of entropy does not yet explain how or why self-organization takes place. Prigogine noted that such self-organization typically takes place in non-linear systems, which are far from their thermodynamic equilibrium state. The thermodynamicists' concrete observations of physical systems were complemented by the more abstract, high level analysis of complex, autonomous systems in cybernetics. The first conference on self-organizing systems, held in 1959 in Chicago, was organized by the same multidisciplinary group of visionary scientists that had founded the discipline of cybernetics.

The British cybernetician W. Ross Ashby proposed what he called "the principle of self-organization". He noted that a dynamical system, independently of its type or composition, always tends to evolve towards a state of equilibrium, or what would now be called an *attractor*. This reduces the uncertainty we have about the system's state, and therefore the system's statistical entropy. This is equivalent to self-organization. The resulting equilibrium can be interpreted as a state where the different parts of the system are mutually adapted.

Another cybernetician, Heinz von Foerster, formulated the principle of "order from noise". He noted that, paradoxically, the larger the random perturbations ("noise") that affect a system, the more quickly it will self-organize (produce "order"). The idea is very simple: the more widely a system is made to move through its state space, the more quickly it will end up in an attractor. If it would just stay in place, no attractor would be reached and no self-organization could take place. Prigogine proposed the related principle of "order through fluctuations". Non-linear systems have in general several attractors. When a system resides in between attractors, it will be in general a chance variation, called "fluctuation" in thermodynamics, that will push it either into the one or the other of the attractors.

2.3. Various applications

Since the 1950's and 1960's, when self-organizing systems were first studied in thermodynamics and cybernetics, many further examples and applications have been discovered. Prigogine generalized his observations to argue for a new scientific world view. Instead of the Newtonian reduction to a static framework ("Being"), he sees the universe as an irreversible "Becoming", which endlessly generates novelty. The cyberneticians went on to apply self-organization to the mechanisms of mind. This gave them a basis for understanding how the brain constructs mental models without relying on outside instruction.

A practical application that grew out of their investigations are the so-called "neural networks", simplified computer models of how the neurons in our brain interact. Unlike the reasoning systems used in artificial intelligence, there is no centralized control in a neural network. All the "neurons" are connected directly or indirectly with each other, but none is in control. Yet, together they manage to make sense out of complex patterns of input.

The same phenomenon, a multitude of initially independent components which end up working together in a coherent manner, appears in the most diverse domains. The production of laser light is a key example of such collective behavior. Atoms or molecules that are excited by an input of energy, emit the surplus energy in the form of photons. Normally, the different photons are emitted at random moments in random directions. The result is ordinary, diffuse light. However, under particular circumstances the molecules can become synchronized, emitting the same photons at the same time in the same direction. The result is an exceptionally coherent and focused beam of light. The German physicist Hermann Haken,

who analysed lasers and similar collective phenomena, was struck by the apparent cooperation or synergy between the components. Therefore, he proposed the new discipline of *synergetics* to study such phenomena.

Another example of spontaneous collective behavior comes from the animal world. Flocks of birds, shoals of fish, swarms of bees or herds of sheep all react in similar ways. When avoiding danger, or changing course, they move together in an elegantly synchronized manner. Sometimes, the swarm or shoal behaves as if it were a single giant animal. Yet, there is no "head fish" or "bird leader", which coordinates the others and tells them how to move. Computer simulations have reproduced the behavior of swarms by letting the individuals interact according to a few simple rules, such as keeping a minimum distance from others, and following the average direction of the neighbours' moves. Out of these local interactions a global, coherent pattern emerges.

2.4. Complex adaptive systems

Swarms are but one of the many self-organizing systems that are now being studied through computer simulation. Whereas before it was very difficult to mathematically model systems with many degrees of freedom, the advent of inexpensive and powerful computers made it possible to construct and explore model systems of various degrees of complexity. This method is at the base of the new domain of "complex adaptive systems", which was pioneered in the 1980's by a number of researchers associated with the Santa Fe Institute in New Mexico. These complexity theorists study systems consisting of many interacting components, which undergo constant change, both autonomously and in interaction with their environment. The behavior of such complex systems is typically unpredictable, yet exhibits various forms of adaptation and self-organization.

A typical example is an ecosystem, consisting of organisms belonging to many different species, which compete or cooperate while interacting with their shared physical environment. Another example is the market, where different producers compete and exchange money and goods with consumers. Although the market is a highly chaotic, non-linear system, it usually reaches an approximate equilibrium in which the many changing and conflicting demands of consumers are all satisfied. The failure of communism has shown that the market is much more effective at organizing the economy than a centrally controlled system. It is as if a mysterious power ensures that goods are produced in the right amounts and distributed to the right places. What Adam Smith, the father of economics, called "the invisible hand" can nowadays simply be called self-organization.

The biologist Stuart Kauffman studied the development of organisms and ecosystems. By computer simulation, he has tried to understand how networks of mutually activating or inhibiting genes can give rise to the differentiation of organs and tissues during embryological development. This led him to investigate the types and numbers of attractors in Boolean networks that represent the pattern of connection between genes. He proposed that the self-organization exhibited by such networks is an essential factor in evolution, complementary to Darwinian selection. Through simulations, he showed that sufficiently complex networks of chemical reactions will necessarily self-organize into autocatalytic cycles, the precursors of life.

Whereas self-organization allows a system to develop autonomously, natural selection is responsible for its adaptation to a variable environment. Such adaptation has been studied most extensively by John Holland, one of the complexity theorists associated with the Santa

Fe Institute. By generalizing from the mechanisms through which biological organisms adapt, he founded the theory of genetic algorithms. This is a general approach to computer problem-solving that relies on the mutation and recombination of partial solutions, and the selective reproduction of the most "fit" new combinations. By allowing the units that undergo variation and selection to interact, through the exchange of signals or "resources", Holland has extended this methodology to model cognitive, ecological and economic systems. These interactions allow simple units to aggregate into complex systems with several hierarchical levels.

Both Holland's and Kauffman's work have provided essential inspiration for the new discipline of *artificial life*. This approach, initiated by Chris Langton, has successfully developed computer programs that mimic lifelike properties, such as reproduction, sexuality, swarming, co-evolution and arms races between predator and prey.

3. Characteristics of self-organizing systems

The different studies which we reviewed have uncovered a number of fundamental traits or "signatures", that distinguish self-organizing systems from the more traditional mechanical systems studied in physics and engineering. Some of these traits, such as the absence of centralized control, are shared by all self-organizing systems, and can therefore be viewed as part of what defines them. Other traits, such as continual adaptation to a changing environment, will only be exhibited by the more complex systems, distinguishing for example an ecosystem from a mere process of crystallization. These traits will now be discussed one by one, starting from the most basic ones, as a prelude to a deeper analysis of the principles of self-organization.

3.1. Two examples: magnetization and Bénard rolls

To make things more concrete, it is useful to keep in mind two basic examples of self-organizing systems. Perhaps the simplest such process that has been extensively studied is magnetization. A piece of potentially magnetic material, such as iron, consists of a multitude of tiny magnets, called "spins" (see Figure 1). Each spin has a particular orientation, corresponding to the direction of its magnetic field. In general, these spins will point in different directions, so that their magnetic fields cancel each other out. This disordered configuration is caused by the random movements of the molecules in the material. The higher the temperature, the stronger these random movements affecting the spins, and the more difficult it will be for any ordered arrangement of spins to maintain or emerge.
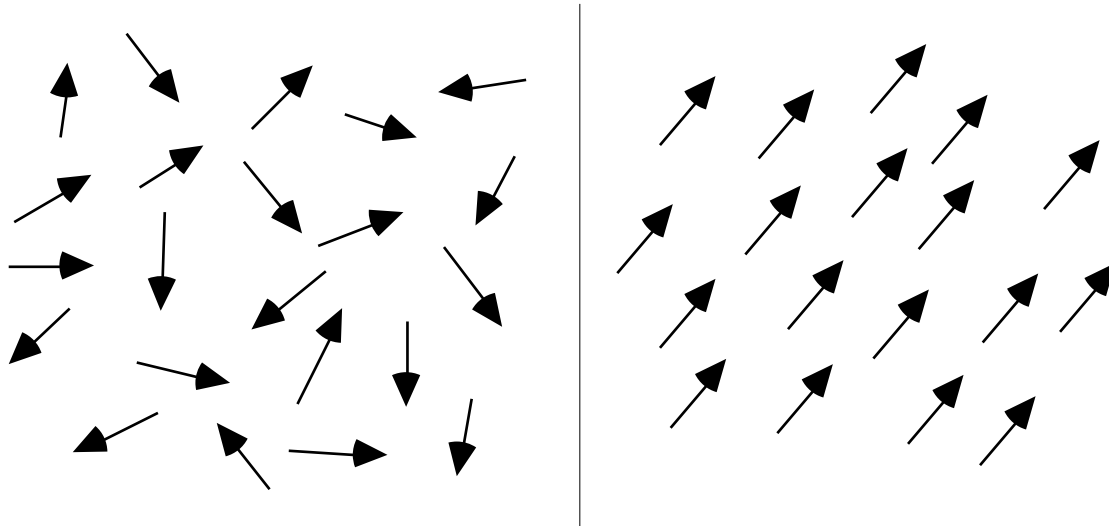
Figure 1: two arrangement of spins: disordered (left) and ordered (right)

However, when the temperature decreases, the spins will spontaneously align themselves, so that they all point in the same direction. Instead of cancelling each other, the different magnetic fields now add up, producing a strong overall field. The reason that the spins "prefer" this ordered arrangement is because spins pointing in opposite directions repel each other, like the North poles of two magnets that are brought together. Spins pointing in the same direction, on the other hand, attract each other, like the North pole of one magnet attracts the South pole of another magnet. Magnetization is a clear case of self-organization, which can be used as a paradigm for a whole range of similar phenomena, such as crystallization (where not only the orientations but also the positions of the molecules become evenly arranged).

A somewhat more complex example will illustrate further characteristics of self-organization. In the Bénard phenomenon, a liquid is heated evenly from below, while cooling down evenly at its surface, like the water in an open container that is put on an electric hot-plate. Since warm liquid is lighter than cold liquid, the heated liquid tries to move upwards towards the surface. However, the cool liquid at the surface similarly tries to sink to the bottom. These two opposite movements cannot take place at the same time without some kind of coordination between the two flows of liquid. The liquid tends to self-organize into a pattern of hexagonal cells, or a series of parallel "rolls", with an upward flow on one side of the roll or cell and a downward flow on the other side. This example is similar to the magnetization in the sense that the molecules in the liquid were moving in random directions at first, but end up all moving in a coordinated way: all "hot" molecules moving upwards on one side of the roll, all "cool" molecules downwards on the other side (see Figure 2). The difference is that the resulting pattern is not static but dynamic: the liquid molecules remain in perpetual movement, whereas the magnetic spins are "frozen" in a particular direction.
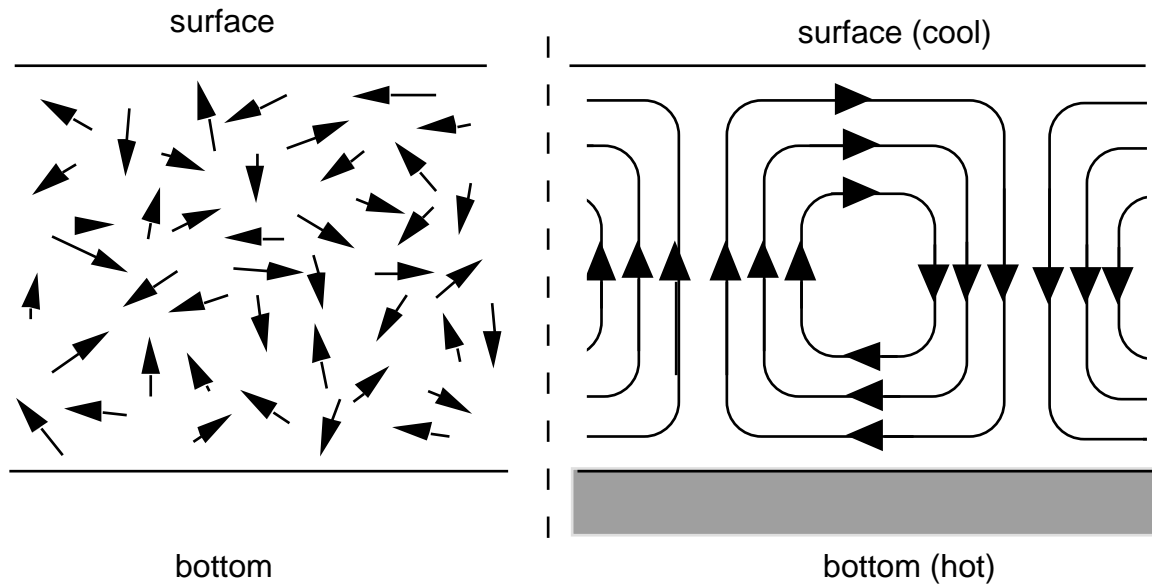
surface | surface (cool)

bottom | bottom (hot)

 Figure 2: two types of movements of liquid molecules: random (left) and in the form of Bénard rolls (right), caused by a difference of temperature between the bottom and the surface of the container.

3.2. Global order from local interactions

The most obvious change that has taken place in both example systems is the emergence of *global* organization. The piece of iron has become magnetic *as a whole*, with a single overall North pole and a single South pole. The liquid as a whole has started cycling through a sequence of rolls. Yet, initially the elements of the system (spins or molecules) were only interacting *locally*. A single spin would only exert a non-negligible influence on its near neighbors. A liquid molecule would only influence the few molecules it collides with.

This locality of interactions follows from the basic continuity of all physical processes: for any influence to pass from one region to another it must first pass through all intermediate regions. During the time that the process propagates through the intervening medium, it will be disturbed by all the fluctuations taking place in that medium. Since we assume that we start with a disordered system, where the components act in random ways, any propagating influence will very quickly be dispersed and eventually destroyed by these random perturbations. As a result, in the original disordered state of the system, distant parts of the system are basically independent: they do not influence each other. Knowing the configuration of the components in one region would give you no information about the configuration in another, non-contiguous region: the configurations have correlation zero.

In the self-organized state, on the other hand, all segments of the system are strongly correlated. This is most clear in the example of the magnet: in the magnetized state, all spins, however far apart, point in the same direction. They have correlation 1. Although the situation is slightly more complicated, the same principle applies to the Bénard rolls: if you know the direction of movement (e.g. up) on one side of the roll, then you can predict the direction not only on the other side (down), but also at the beginning of the next roll, the roll after that, and so on. Since the rolls are evenly spaced, this means that if you know the

diameter of a roll, you can predict the movement at any distance from the point you started with.

Correlation is a useful measure to study the transition from the disordered to the ordered state. Locality implies that neighboring configurations are strongly correlated, but that this correlation diminishes as the distance between configurations increases. The *correlation length* can be defined as the maximum distance over which there is a significant correlation. For the magnet, in the disordered state the correlation length is practically zero, in the ordered state it spans the complete length of the magnet. In the intermediate stages, while the magnet is self-organizing, the correlation length gradually increases.

The mechanism can be pictured as follows. Consider the material at a temperature where the spins still move randomly, but the movement is not so strong that it can destroy tightly coupled configurations. By chance, some of these random movements will result in two or three neighboring spins pointing in the same direction. The magnetic fields of these spins now reinforce each other, and therefore will have a stronger influence on their neighbors than the non-aligned fields. If one of the neigbors by another fluctuation happens to align itself with the group, the magnetic force will make it more difficult for further fluctuation to dislodge it from that position. The more spins align themselves with the initial assembly, the stronger the overall force they exert on their neighbors, and the more likely it will be that these neighbors too align themselves to their orientation. Thus, the overall alignment will propagate like a concentric wave, extending ever further from the initial assembly, until all spins are aligned. The local alignment has expanded into a global order.

3.3. Distributed control

When we consider a highly organized system, we usually imagine some external or internal agent that is responsible for guiding, directing or controlling that organization. For example, most human organizations have a president, chief executive or board of directors that develops the policies and coordinates the different departments. The actions of our body are largely planned and controlled by our brain. The activity of a cell is largely determined by the "blueprint" stored in its chromosomes. In each of these cases, although the controlling agent (president, brain or chromosome) is part of the system, it is in principle possible to separate it from the rest. The controller is a physically distinct subsystem, that exerts its influence over the rest of the system. In this case, we may say that control is *centralized*.

In self-organizing systems, on the other hand, "control" of the organization is typically *distributed* over the whole of the system. All parts contribute evenly to the resulting arrangement. In the example of the magnet, it may seem that the initial assembly that triggered the process is more important than the others. Yet, its role could have been played by any other group of spins that happened to align itself by chance. In practice, different alignments will appear independently in different parts of the material, and compete for the recruitment of the remaining non-aligned spins. That competition is usually won by the assembly that has grown largest, but the system may also settle into a number of differently aligned zones. (in this case, the correlation length will be less than maximal). Once the alignment has spread, no spin or group of spins can change that configuration. The triggering assembly does not have any more power in that respect than others. The overall magnetization is kept up because all spins work together: the "responsibility" is shared. That is why it is so difficult for any subsystem to deviate from, or influence, that pattern.

Although centralized control does have some advantages over distributed control (e.g. it allows more autonomy and stronger specialization for the controller), at some level it must itself be based on distributed control. For example, the behavior of our body can be best explained by studying what happens in our brain, since the brain, through the nervous system, controls the movement of our muscles. However, to explain the functioning of our brain, we can no longer rely on some "mind within the mind" that tells the different brain neurons what to do. This is the traditional philosophical problem of the *homunculus*, the hypothetical "little man" that had to be postulated as the one that makes all the decisions within our mental system. Any explanation for organization that relies on some separate control, plan or blueprint must also explain where that control comes from, otherwise it is not really an explanation. The only way to avoid falling into the trap of an infinite regress (the mind within the mind within the mind within...) is to uncover a mechanism of self-organization at some level.

The brain illustrates this principle nicely. Its organization is distributed over a network of interacting neurons. Although different brain regions are specialized for different tasks, no neuron or group of neurons has overall control. This is shown by the fact that minor brain lesions because of accidents, surgery or tumors normally do not disturb overall functioning, whatever the region that is destroyed.

3.4. Robustness, resilience

The same effect is simulated on computers by neural networks. A neural network that has been "trained" to perform a certain task (e.g. recognize hand-written letters) will in general still be able to perform that task when it is damaged, for example by the random removal of nodes and links from the network. Increasing damage will decrease performance, but the degradation will be "graceful": the quality of the output will diminish gradually, without sudden loss of function. A traditional computer program or mechanical system, on the other hand, will in general stop functioning if random components are removed.

This is a general characteristic of self-organizing systems: they are *robust* or *resilient*. This means that they are relatively insensitive to perturbations or errors, and have a strong capacity to restore themselves, unlike most human designed systems. For example, an ecosystem that has undergone severe damage, such as a fire, will in general recover relatively quickly. In the magnet, if part of the spins are knocked out of their alignment, the magnetic field produced by the rest of the spins will quickly pull them back.

One reason for this fault-tolerance is the redundant, distributed organization: the non-damaged regions can usually make up for the damaged ones. This can be illustrated by holography: unlike a normal photograph, the information about an element of the pictured scene is distributed over the whole of the hologram. Damaging or cutting out part of the holographic sheet will merely make the resulting image more fuzzy, without removing any essential component of the picture.

Another reason for this intrinsic robustness is that self-organization thrives on randomness, fluctuations or "noise". Without the initial random movements, the spins would never have discovered an aligned configuration. The same applies to the molecules in the Bénard rolls. It is this intrinsic variability or diversity that makes self-organization possible. As will be discussed later, a certain amount of random perturbations will facilitate rather than hinder self-organization. A third reason for resilience, the stabilizing effect of feedback loops, will be discussed in the next section.

3.5. Non-linearity and feedback

Most of the systems modelled by the traditional mathematical methods of physics are *linear.* This means basically that effects are proportional to their causes: if you kick a ball twice as hard, it will fly away twice as fast. In self-organizing systems, on the other hand, the relation between cause and effect is much less straightforward: small causes can have large effects, and large causes can have small effects. Imagine that you subject a magnetized piece of iron to an external magnetic field with a direction different from its own field. If the external field is relatively small, it will have no effect: the internally generated field is strong enough to keep the spins aligned, independently of the outside field. The magnetization is robust. Suppose that the external field is gradually increased. At a certain point it will become stronger than the magnet's own field. At that point, thermal fluctuations that make a spin move parallel to the outside field rather than to its neighbors will no longer be counteracted, and all the spins will quickly start to align themselves with the external field. In the beginning, a large increase in the external field has practically no effect, until a threshold is crossed and a small further increase suddenly turns the whole system around.

This non-linearity can be understood from the relation of *feedback* that holds between the system's components. Each component (e.g. a spin) affects the other components, but these components in turn affect the first component. Thus the cause-and-effect relation is *circular*: any change in the first component is fed back via its effects on the other components to the first component itself. Feedback can have two basic values: *positive* or *negative.* Feedback is said to be positive if the recurrent influence reinforces or amplifies the initial change. In other words, if a change takes place in a particular direction, the reaction being fed back takes place in that same direction. Feedback is negative if the reaction is opposite to the initial action, that is, if change is suppressed or counteracted, rather than reinforced. Negative feedback stabilizes the system, by bringing deviations back to their original state. Positive feedback, on the other hand, makes deviations grow in a runaway, explosive manner. It leads to accelerated development, resulting in a radically different configuration.

A process of self-organization, such as magnetization or the emergence of Bénard rolls, typically starts with a positive feedback phase, where an initial fluctuation is amplified, spreading ever more quickly, until it affects the complete system. Once all components have "aligned" themselves with the configuration created by the initial fluctuation, the configuration stops growing: it has "exhausted" the available resources. Now the system has reached an equilibrium (or at least a stationary state). Since further growth is no longer possible, the only possible changes are those that reduce the dominant configuration. However, as soon as some components deviate from this configuration, the same forces that reinforced that configuration will suppress the deviation, bringing the system back to its stable configuration. This is the phase of negative feedback.

In more complex self-organizing systems, there will be several interlocking positive and negative feedback loops, so that changes in some directions are amplified while changes in other directions are suppressed. This can lead to very complicated, difficult to predict behavior.

3.6. Organizational closure, hierarchy and emergence

The correlation or coherence between separate components produced by self-organization defines an ordered configuration. However, order does not yet mean organization. Organization can be defined as the characteristic of being ordered or structured so as to fulfil a particular *function*. In self-organizing systems, this function is the maintenance of a particular configuration, in spite of disturbances. Only those orders will result from self-organization that can maintain themselves. This general characteristic of self-sufficiency can be understood through the concept of *closure*.

A general causal process can be analysed as a chain or sequence A   B   C   D   ... of situations or events, such that the first event A causes the next event B, B causes C, and so on. In general, this produces ongoing change. However, it is possible that at some stage the chain closes in on itself, so that O leads back to an earlier stage J. In that case, the system will continue to cycle through J, K, L, M, N, O, J, K, L, ... The corresponding arrangement of the system will be continuously maintained or reproduced. If the cycle moreover settles into a negative feedback regime, it will be relatively impervious to external disturbances. The system has now become responsible for its own maintenance, and thus become largely independent from the environment. It is thus also "closed" against influences from the outside. Although in general there will still be exchange of matter and energy between system and environment, the organization is determined purely internally. Thus the system is thermodynamically open, but organizationally closed.

This closed nature of natural organizations is illustrated by the cyclical flow in a Bénard roll. However, as discussed in the previous section, this circular causality also underlies magnetization, albeit in a more diffuse manner. For the outside observer, closure determines a clear distinction between inside (the components that participate in the closure) and outside (those that do not), and therefore a boundary separating system from environment. This boundary can encompass all components of the original system, like in magnetization, or only part of them, as in a Bénard roll. This defines a single Bénard roll as a new subsystem, which is clearly distinguishable from the other rolls. An individual roll or cell will exchange only little material or energy with its neighboring rolls, and thus can be viewed as a largely autonomous system. The same partitioning can appear during magnetization if different regions settle into different spin alignments.

More generally, a self-organizing system may settle into a number of relatively autonomous, organizationally closed subsystems, but these subsystems will continue to interact in a more indirect way. These interactions too will tend to settle into self-sufficient, "closed" configurations, determining subsystems at a higher hierarchical level, which contain the original subsystems as components. These higher level systems themselves may interact until they hit on a closed pattern of interactions, thus defining a system of a yet higher order. This explains why complex systems tend to have a hierarchical, "boxes within boxes" architecture, where at each level you can distinguish a number of relatively autonomous, closed organizations. For example, a cell is an organizationally closed system, encompassing a complex network of interacting chemical cycles within a membrane that protects them from external disturbances. However, cells are themselves organized in circuits and tissues that together form a multicellular organism. These organisms themselves are connected by a multitude of cyclical food webs, collectively forming an ecosystem.

Organizational closure turns a collection of interacting elements into an individual, coherent whole. This whole has properties that arise out of its organization, and that cannot be reduced to the properties of its elements. Such properties are called *emergent*. For example, a Bénard

roll is characterized by the direction of its rotation. For an independent molecule, such rotation is not defined. A higher level, emergent property will typically constrain the behavior of the lower level components. For example, the overall rotation characterizing a Bénard roll will force the liquid molecules to move in particular directions rather than others. This constraint cannot be understood from the interactions on the level of a molecule: it is determined by the emergent level. This is called *downward causation*: it is as if the higher level exerts its influence downward to the lower level, causing the molecules to act in a particular way. Downward causation is to be contrasted with the more traditional "upward" causation underlying Newtonian reductionism, where the behavior of the whole is fully determined by the behavior of the parts.

3.7. Bifurcations, symmetry breaking

Linear systems of equations normally have a single solution. Non-linear systems, on the other hand, have typically several solutions, and there is no a priori way to decide which solution is the "right" one. In terms of actual self-organizing systems, this means that there is a range of stable configurations in which the system may settle. This is obvious in the case of magnetization: there is no preference for any one direction of spin alignment. Each direction of the resulting magnetic field is an equally likely outcome of self-organization. For Bénard rolls, if we assume that the container has a fixed shape, there are two possible stable configurations: either a given roll rotates clockwise or it rotates counter clockwise.

Which of the possible configurations the system will settle in will depend on a chance fluctuation. Since small fluctuations are amplified by positive feedback, this means that the initial fluctuation that led to one outcome rather than another may be so small that it cannot be observed. In practice, given the observable state of the system at the beginning of the process, the outcome is therefore *unpredictable*.

However, if we go back to the state of the system before self-organization, there was only one possible configuration: a disordered one. A disordered configuration is one in which the possible states for the individual components have the same probability: all directions are equally likely for a spin, or for the movement of a liquid molecule. Since the number of components is typically astronomical, it follows from the law of large numbers that the different microscopic states or directions will cancel each other out, so that the overall magnetization or liquid flow is zero. Although the individual components all behave differently, on the global, macroscopic level, the system is homogeneous: every direction of spin or movement is represented to the same degree. This means that the system is *symmetric*: from whatever direction it is observed, it will look the same.

After self-organization, however, one direction or one configuration dominates all others, and therefore the symmetry is lost. The angle of magnetization changes if the magnet is observed from a different point of view. The Bénard rolls looks differently if the viewpoint rotates 180 degrees: what was clockwise becomes counter clockwise. Another way to conceive of such *symmetry breaking* is to look at the self-organizing system as if it were making a choice: initially it treated all configurations equally, but then it expressed a preference for one possibility. However, there are no objective criteria for preferring one stable configuration over another one. It is as if the system has made an arbitrary decision, and thus changed the range of possibilities. It is this unpredictability which, in a sense, creates the real novelty.

The evolution from disordered to ordered configuration is usually triggered by a change in the external situation, the *boundary conditions* of the system. For example, in the Bénard phenomenon, rolls will only appear if the temperature difference between bottom and surface becomes large enough. For smaller differences, heat will be exchanged by mere diffusion, and there will not be any permanent flows in the liquid. Let us consider the average speed of the molecules in a particular region of the liquid. Before the transition, that speed is zero. After the appearance of the rolls, the molecules in that region will all move down, or all up, depending on whether the local roll settles into the clockwise or the counter clockwise regime. In general, the higher the temperature difference, the larger the speed. This is best visualized through a diagram (Figure 3) where the speed in the region is expressed as a function of the temperature difference. This diagram shows a clear branching or *bifurcation* in the possible states as the temperature difference increases: at a particular value t_0, the zero speed configuration becomes unstable, and is replaced by two different stable configurations, up and down.
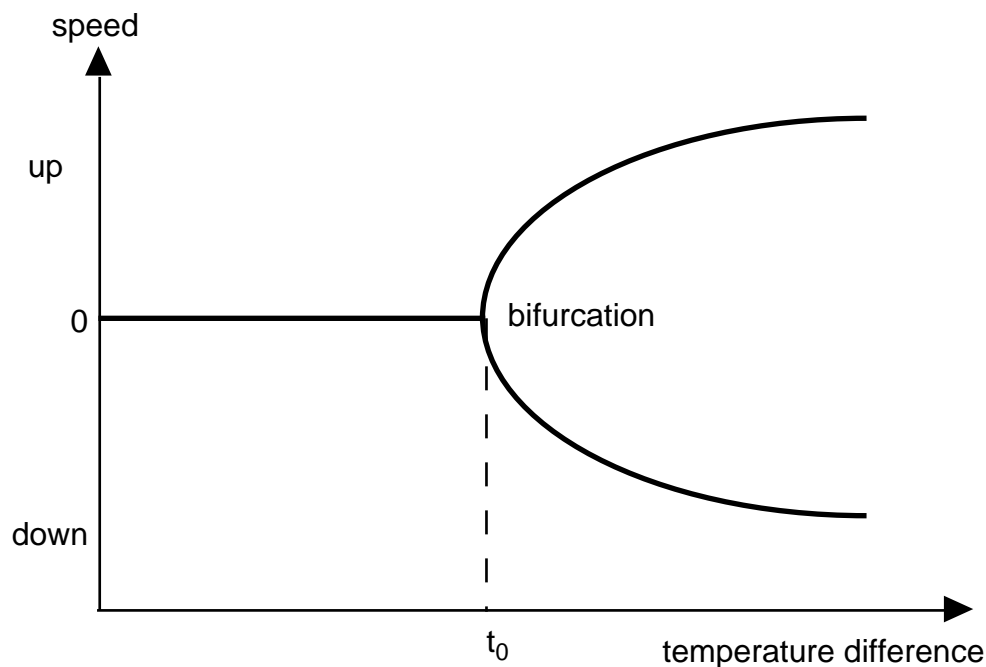

Figure 3: the bifurcation characterizing the appearance of Bénard rolls: when the temperature difference increases, at a certain point t_0, there are two possible outcomes for the average speed of the liquid molecules; instead of speed zero (no movement), the molecules either move upwards or downwards.

Many self-organizing processes can be described by this type of diagram, where the types of possible solutions are depicted in function of a certain variable whose values determine the onset of self-organization. Since it marks the transition between ordered and disordered configurations, this variable is usually called an *order parameter*. Much more complicated bifurcations are possible: instead of two, there can be three, four or any number of solutions appearing at a bifurcation point, and bifurcations may be arranged in a "cascade", where each branch of the fork itself bifurcates further and further as the order parameter increases. In some cases, bifurcations appear ever more quickly as the order parameter increases, until the

number of branches becomes infinite. This typically characterizes the onset of the chaotic regime, where the system jumps constantly and unpredictably from one branch (configuration) to another one.

As is the case with the temperature difference, the order parameter often measures the distance from thermodynamic equilibrium, i.e. the minimum energy state where there is no activity. The increase in the number of possible configurations that accompanies the increase in the order parameter can be seen as an increase in general variability or "turbulence" that distance from equilibrium engenders. One way to understand this is by noting that more energy "pumped" into the system allows more amplification of small differences (positive feedback), and therefore more varied types of behavior. Let us study this in more detail.

3.8. Far-from-equilibrium dynamics

The difference between magnetization and the Bénard phenomenon is that the former ends up in a static state of *equilibrium*, whereas the latter produces a *stationary state* of on-going activity. In thermodynamics, equilibrium is characterized by the absence of entropy production, or, equivalently, by the fact that no energy is dissipated. A system in equilibrium has settled in a minimum of its potential energy function. To reach that state it had to dissipate all the "surplus" energy it contained. Without external input of energy, it will remain forever fixed in that minimal energy state. In the Bénard case, on the other hand, the heating of the liquid from below provides a constant influx of energy. Therefore, the system cannot reach equilibrium. At best, it can get rid of the surplus energy that enters the system by dissipating it into the cooler environment above. Prigogine and others have suggested that in such cases the second law of thermodynamics, which states that the total entropy in a closed system reaches a maximum, should be replaced by a new law of *maximum entropy production*: in systems far from thermodynamic equilibrium the dissipation of entropy to the environment reaches a maximum. This law remains controversial, though.

The constant input of energy in far-from-equilibrium systems entails continuous movement or *flow* among the components of the system. This flow will be fed by low entropy matter or energy entering from the environment, which will cycle through the system, undergoing a number of conversions, in order to finally leave the system as a high entropy output. For example, an ecosystem thrives on an input of sunlight and minerals, which are transformed by plants into organic matter, which is itself converted via a number of intermediary stages by animals, bacteria and fungi back into minerals and heat. Although the minerals and related resources such as water, oxygen and carbon dioxide are constantly recycled, the energy of the sun is dissipated and becomes effectively unusable. Similarly, an economy uses raw materials and energy (e.g. petroleum) in order to produce goods and waste products (e.g. carbon dioxide and heat).

The dependency on an external source of energy makes a far-from-equilibrium system more fragile and sensitive to changes in the environment, but also more dynamic and capable to react. The fragility is obvious: if the energy source were to disappear (e.g. if the heating under the Bénard rolls would be switched off), the dissipative structure would disintegrate. On the other hand, the surplus of energy allows the system to amplify on-going processes, e.g. counteracting small perturbations by large reactions, or sustaining positive feedback cycles for a much longer time. This makes the system much more powerful in developing, growing or adapting to external changes. Instead of reacting to all perturbations by a negative feedback that brings the system back to the same state of equilibrium, a far-from-equilibrium system is in principle capable of producing a much greater variety of regulating actions, leading to

multiple stable configurations. In order to maintain a particular organization in spite of environmental change, though, the question is which action to use in which circumstances. This defines the problem of adaptation.

4. Characteristics of adaptive systems

4.1. Adaptation as fit

A configuration of a system may be called "fit" if it is able to maintain or grow given the specific configuration of its environment. An unfit configuration, on the other hand, is one that will spontaneously disintegrate under the given boundary conditions. Different configurations can be compared as to their degree of fitness, or likeliness to survive under the given conditions imposed by the environment. Thus, adaptation can be conceived as achieving a *fit* between system and environment.

It follows that every self-organizing system adapts to its environment. The particular stable configuration it reaches by definition fits its particular circumstances. For example, the pattern and speed of flow in the Bénard rolls will be adapted to the specific temperature difference between bottom and surface, whereas the orientation of the spins will tend to be parallel to any outside magnetic field. In this sense, self-organization implies adaptation. This becomes even more clear if we choose a different boundary to distinguish system from environment. As noted by Ashby, if we consider a particular part of the original, self-organized system as the new "system", and the remainder as its "environment", then the part will be necessarily adapted to the environment. For example, in the magnet the orientation of a particular segment of spins will be adjusted to the magnetic field generated by the rest of the spins.

For a given boundary, adaptation becomes less trivial when the boundary conditions change. For large changes, this in general means that the existing configuration becomes unstable. This may lead to its disintegration, and the need to start the process of self-organization anew. This may not seem a great problem for systems like the magnet or the Bénard rolls, but it would be disastrous for more complex systems such as organisms, ecosystems or societies. Systems may be called *adaptive* if they can adjust to such changes while keeping their organization as much as possible intact.

4.2. Regulation and the edge of chaos

Cybernetics has shown that adaptation can be modelled as a problem of *regulation* or *control*: minimizing deviations from a goal configuration by counteracting perturbations before they become large enough to endanger the essential organization. This means that the system must be able to: 1) produce a sufficient *variety* of actions to cope with each of the possible perturbations (Ashby's "law of requisite variety"); 2) *select* the most adequate counteraction for a given perturbation. Mechanical control systems, such as a thermostat or an automatic pilot, have both variety and selectivity built in by the system designer. Self-organizing systems need to autonomously evolve these capabilities.

Variety can be fostered by keeping the system sufficiently far from equilibrium so that it has plenty of stationary states to choose from. Selectivity requires that these configurations be sufficiently small in number and sufficiently stable to allow an appropriate one to be "chosen" without danger of losing the overall organization. This explains the observation,

made by researchers such as Langton and Kauffman, that complex adaptive systems tend to reside on the "edge of chaos", that is, in the narrow domain between frozen constancy (equilibrium) and turbulent, chaotic activity. The mechanism by which complex systems tend to maintain on this critical edge has been called *self-organized criticality* by Per Bak. The system's behavior on this edge is typically governed by a "power law": large adjustments are possible, but are much less probable than small adjustments.

Limiting the degree of chaos in the system's activities is not yet sufficient for the selection of appropriate actions, though. The system needs a fitness criterion for choosing the best action for the given circumstances. The most straightforward method is to let the environment itself determine what is fit: if the action maintains the basic organization, it is, otherwise it is not. This can be dangerous, though, since trying out an inadequate action may lead to the destruction of the system. Therefore, complex systems such as organisms or minds have evolved internal *models* of the environment. This allows them to try out a potential action "virtually", in the model, and use the model to decide on its fitness. The model functions as a *vicarious selector*, which internally selects actions acting for, or in anticipation of, external selection. This "shortcut" makes the selection of actions much more reliable and efficient. It must be noted, though, that these models themselves at some stage must have evolved to fit the real environment; otherwise they cannot offer any reliable guidance. Usually, such models are embodied in a separate subsystem, such as the genome or the brain, and thus they would rather fit the paradigm of centralized control than the one of self-organization. Therefore, we will not further develop this complex, but important subject.

4.3. Variation and selection

Neither the magnet nor the Bénard rolls are adaptive in this more advanced sense. Therefore, it is worth mentioning a few other examples of complex adaptive systems. The immune system maintains a living system's organization by destroying parasitic bodies, such as bacteria or cancer cells. It achieves this by producing antibodies that attach themselves to the alien bodies and thus neutralize them. To find the right type of antibodies, the immune system simply produces an astronomical variety of different antibody shapes. However, only the ones that "fit" the invaders are selected and reproduced in large quantities.

A similar type of simple variation and selective production of fit components underlies economic and ecological adaptation mechanisms. When the circumstances in an ecosystem change (e.g. the climate becomes drier), those species or varieties in the ecosystem that are better adapted will become more numerous relative to those that are less, until the balance is restored. Similarly, if in an economy a particular good becomes scarce (e.g. because of increased demand, or decreased supply in the resources needed to produce that good), the firms producing that good will get higher prices for their product, and thus will be able to invest more in its production. This will allow them to grow relative to other firms, again until the balance is restored.

Both examples assume that the initial variety in species or in firms is sufficient to cope with the perturbation. However, if none of the existing species or firms would be really fit to cope with the new situation, new variations would arise, e.g. by mutation or sexual recombination in the case of species, and by R&D efforts or new start-ups in the case of firms. If the system is sufficiently rich in inherent diversity and capacity to evolve, this variation will sooner or later produce one or more types of component that can "neutralize" the perturbation, and thus save the global system.

In the above description, fit components are selected directly by the environment. In practice, only some components will adapt to the external changes, but the resulting internal changes in the system will trigger a further round of adaptations, where a change in one component will create a change in the other components that interact with it, which in turn will change the components they interact with, and so on, until some kind of equilibrium has been restored. For example, the climate becoming dryer in a semi-tropical forest will lead to a an increase in drought-resistant trees, such as Eucalyptus. This in turn will lead to an increase in the population of animals that feed on Eucalyptus leaves, and a decrease in the populations that cannot digest Eucalyptus leaves. The growth in a Eucalyptus eating population, such as koalas, will create opportunities for species that depend on koalas, such as species-specific fleas or other parasites. A similar effect of environmental changes gradually rippling through the system can be found in the economy: the appearance or growth of a particular type of producer will boost the growth of firms supplying raw materials or services to that kind of producer, and so on.

A very simple analogy of this effect can even be found in the magnet: an external field will be felt first of all by the spins on the outer fringe of the magnet, where the field created internally by the other spins is weakest. These spins will be the first to align themselves to the outside field, followed by the next, more interior layer, and so on, until even the inner core has switched orientation.

5. The state-space description of self-organization

After this rather intuitive overview of the properties characterizing self-organizing systems, it is time to present some of the more abstract, formal concepts and principles that allow us to model self-organization and adaptation in the most general way. The modelling approach that will be introduced here can be used to build precise mathematical models or computer simulations for specific systems. However, since the variety and complexity of such models is unlimited, and since the more typical examples are discussed in depth in the literature, the present overview will focus merely on the general principles that these models have in common.

5.1. State spaces

Every system capable of change has certain variable features, that can take on different values. For example, a particle can have different positions and move with different speeds, a spin can point in different directions, and a liquid can have different temperatures. If such a feature can be described as a real number varying over a finite or infinite interval, it is called a degree of freedom. All the values for the different variables we consider together determine the *state s* of the system.

Any change or evolution of the system can be described as a transition from one state to another one. When defining the state, we normally try to include all the variables whose values are relevant for the further evolution of the system, but not more. For example, when we wish to predict the movement of a billiard ball, we will consider its position, momentum and possibly rotation, but not its color, as part of its state.

The set of all possible states of a system is called its state space: $S = \{s_1, s_2, s_3, ...\}$. If all relevant variables can be described by degrees of freedom, their number determines the dimensionality of the space. However, the concept of state space is more general than the one

of a continuous, dimensional space, and can include any set with finite, discrete or continuous number of elements. For example, in a 1-dimensional case, a spin can have only two possible states: up or down. This can be represented by a binary, Boolean variable, with the values 1 ("up") and 0 ("down"). For simplicity, we will assume from now on that the state space is discrete, and even finite, although all mathematical expressions can be generalized to the continuous case.

If a system A consists of n different subsystems or components A^1, A^2, A^3, ..., A^n that can vary independently, then A's overall state space $S(A)$ is the Cartesian product of the state spaces of its components: $S = S^1 \times S^2 \times ... S^n$, s $S = (s^1, s^2, ..., s^n)$. The dimension (number of degrees of freedom) of S is the product of the dimensions of all of the component spaces. Self-organizing systems usually consist of a very large number of components (in thermodynamics, n is typically of the order of the Avogadro number, 10^{23}). This means that the state spaces are so huge that no existing mathematical or computational technique is capable of calculating or otherwise determining the exact state. It is also impossible to determine the state by observation, since this means that all determining variables would need to be measured for each of the 10^{23} components!

Therefore, self-organizing systems can only be modelled by statistical methods, that is by calculating the *probability* $P(s)$ that the system is in a particular state s, given a limited number of properties that have been determined by observation. The function P: S $[0, 1]$: s $P(s)$ attaches a probability to each state and thus determines a probability distribution over the state space.

5.2. Uncertainty and entropy

When nothing is known about the system, there is no reason to assume that one state is more likely than another state, and therefore all states have the same probability: $P(s) = P(s')$, s, s' S. The probability distribution is homogeneous. On the other hand, having full information means that we have certainty that the system is in a particular state s: $P(s) = 1$, $P(s') = 0$, s' s S.

More generally, we can use Shannon's theory of information to determine our degree of uncertainty H about the system:

$$H(P) = - \sum_{s \in S} P(s).\log P(s)$$

(if the state space is continuous, the sum can be replaced by an integral.) It is easily seen that uncertainty is minimal ($H = 0$) when one state has probability 1 and all others have probability 0, and maximal ($H = \log N$, with N the number of states in S), when all states have the same probability. When we are certain that the state resides within a subspace S_0 S, but do not know anything more, uncertainty takes an intermediate value: $H = \log N_0$, with $1 < N_0 < N$ the number of states in S_0. The information I we receive from an observation is equal to the degree to which uncertainty is reduced by the observation: $I = H$(before) - H(after).

Shannon's expression for uncertainty appears to be the same, apart from a constant factor, as the expression that Boltzmann proposed for the entropy of a system in statistical mechanics.

Boltzmann's statistical entropy, which measures the degree of disorder, seems in many, but not all, cases to behave the same as the original thermodynamic entropy S, which is defined by the degree of dissipation of heat. These apparent equivalencies have led to many interesting, but confusing, speculations about the nature of disorder, entropy and self-organization. To minimize confusion, we will from now on always consider Shannon's uncertainty and Boltzmann's statistical entropy as identical, and only speak about thermodynamic entropy insofar that it behaves differently from this basic concept.

H can in general be interpreted as a measure of our ignorance about the system's state, or, alternatively, as a measure of the freedom the system still has in "choosing" a state, given the information we have. Reducing H can be interpreted similarly as either gaining information, or putting a *constraint* on the system, so as to restrict its freedom of choosing a state.

Self-organization as the appearance of coherence or correlation between the system's components is equivalent to the reduction of H. Indeed, if the state of one component becomes dependent on another component's state, then it loses its freedom to vary. The global order imposed on the system means that the overwhelming majority of states in the system's state space are no longer available. For example, in the case of the magnet, each spin has 2 degrees of freedom (spherical coordinates) along which its orientation can vary. For a system consisting of N spins, there are 2N degrees of freedom, an astronomical number. After self-organization, however, all spins are aligned to point in the same direction, and there are only 2 degrees of freedom left.

5.3. Attractors

To model the evolution of a system, we need rules that tell us how the system moves from one state to another in the course of time t. This might be expressed by a function f_T: $S \quad S$: $s(t) \quad s(t+T)$, which is usually the solution of a differential or difference equation. According to classical mechanics, the evolution of a system is deterministic and reversible. In practice, though, the evolution of complex systems is *irreversible*: the future is fundamentally different from the past, and it is impossible to reconstruct the past from the present. This follows from the fact that self-organizing systems dissipate energy: dissipated energy cannot be recovered in order to undo the process.

Most models still assume the dynamics to be deterministic, though: for a given initial state s, there will in general be only one possible later state $f(s)$. In practice, the lack of information about the precise state will make the evolution unpredictable. The corresponding stochastic process can be modelled as a Markov chain, which for each initial state s_i gives the probability that the system will undergo a transition to a next state s_j: $P(s_j|s_i) = \mathbf{M}_{ij} \quad [0, 1]$. $\mathbf{M}$ is the transition matrix of the Markov chain. Given a probability distribution $P(s_i, t)$ for the initial state, this makes it possible to calculate the probability distribution for the next state:

$$P(s_j, t+1) = \sum_i \mathbf{M}_{ij} P(s_i, t)$$

Both deterministic and stochastic evolutions will in general be *asymmetric* in time, that is, there will be transitions between states: $s_i \quad s_f$, such that the inverse transition $s_f \quad s_i$ is either impossible or less probable. For example a ball on top of a mountain (state s_i) will tend to roll down to the bottom of a valley (state s_f), but the reverse is impossible or at least very

unlikely. This means that when a dynamic or stochastic system is allowed to evolve, it will tend to leave certain states (e.g. s_i) and to enter and stay in other states (e.g. s_f).

This intuition can be expressed more precisely by introducing the concept of an *attractor*. An attractor is a subset A of the system's state space S that the system can enter but not leave, and which contains no smaller subset with the same property. This means that s_i A, s_j A, $n, T: f_T(s_i)$ A or $M^n_{ij} = 0$. The property of not containing a smaller such set can be expressed as: $\lim_n$ $M^n_{ik} = 0$ if and only if s_k A.

An attractor is a mathematical model of causal closure: inside the attractor, the process has "closed in" on itself, and can no longer reach out. Attractors can have many different shapes, sizes and dimensions. The simplest one is a zero-dimensional point attractor which consists of a single state. This describes the situation where a system reaches an equilibrium. Very common is also a one-dimensional limit cycle, where all states of the attractor are revisited at regular intervals. This describes certain far-from-equilibrium configurations where the system exhibits periodical behavior, such as the Bénard rolls. A so-called "strange" attractor is characterized by a non-integer, fractal dimension. This is characteristic of certain chaotic processes.

Attractors in general have a *basin* $B(A)$: a set of states outside a given attractor whose evolution necessarily ends up inside: s $B(A)$: s A and T such that $f_T(s)$ A. In a deterministic system, every state either belongs to an attractor or to an attractor basin. In a stochastic system, there is a third category of states that can end up in either of several attractors.

When a system enters an attractor, it thereby loses its freedom to reach states outside the attractor, and thus decreases its statistical entropy H. During the system's "descent" from the basin into the attractor, the entropy will gradually decrease, reaching a minimum in the attractor itself. The remaining entropy will depend on the size of the attractor, in the simplest case: $H = \log N$ with $N = \#(A)$. Since the reaching of an attractor is an automatic process, entailed by the system's dynamics, it can be viewed as a general model of self-organization: the spontaneous reduction of statistical entropy.

5.4. Fitness landscapes

In many cases, the rather abstract and mathematically complex structure of a system of attractors and basins can be replaced by the more intuitive model of a fitness landscape. Under certain mathematical conditions, the deterministic dynamics of a system can be represented by a *potential* function F on the state space, which attaches a certain number to each state: $F: S$ $\mathbf{R}: s$ $F(s)$, such that trajectory of the system through the state space will always follow the path of steepest descent, i.e. move from a given state s to that neighboring state for which F is minimal. In mechanics, this function corresponds to the potential energy of the system. More generally, the function represents the degree to which a certain state is "preferable" to another state: the lower the value of F, the "better" or the more "fit" the state. Thus, the potential can be seen as the negative of the fitness function. It is unfortunate that the convention in physics sees systems as striving to minimize a potential function, whereas the convention in biology sees systems as striving to maximize a fitness function. Although this tends to be confusing, the two types of representation are equivalent apart from an inversion of the sign of the function.

The fitness function transforms the state space into a *fitness landscape*, where every point in the space has a certain "height" corresponding to its fitness value. This landscape has peaks and valleys. The attractors of the dynamics will now correspond to the local minima of the potential function (potential valleys), or, equivalently, to the maxima of the fitness function (fitness peaks). This can be understood by noting that the system will always move downward in the potential landscape. When it has reached the locally lowest point, all remaining directions will point upward, and therefore the system will not be able to leave the bottom of the valley (see figure).

The local maxima of the potential (peaks) are the points that separate the basins of the attractors (valleys) that lie in between the peaks. In general, the steeper the slope, the faster the descent of the system along that slope.
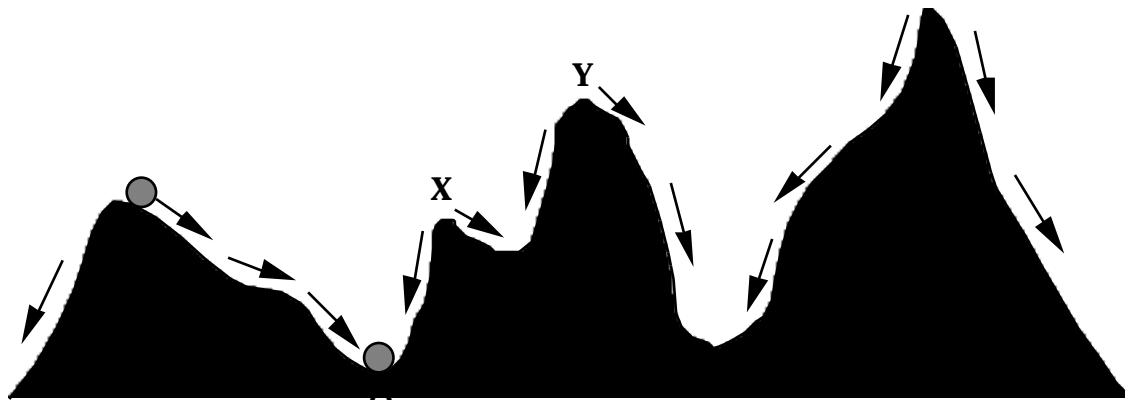

Figure 4: a fitness landscape: the arrows denote the directions in which the system will move. The height of a position corresponds to its potential or to its lack of fitness. Thus, A has a higher fitness (or lower potential) than B. The bottoms of the valleys A, B and C are local minima of the potential, i.e. attractors. The peaks X and Y delimit the basin of the attractor B. X separates the basins of A and B, Y separates the basins of B and C.

5.5. Order from noise

In the fitness landscape representation all attractors are not equal: those with a higher fitness are in a sense "better" than the others. For self-organizing systems, "better" or "fitter" usually means "more stable" or "with more potential for growth". However, the dynamics implied by a fitness landscape does not in general lead to the overall fittest state: the system has no choice but to follow the path of steepest descent. This path will in general end in a *local* minimum of the potential, not in the *global* minimum.

Apart from changing the fitness function, the only way to get the system out of a local minimum is to add a degree of indeterminism to the dynamics, that is, to give the system the possibility to make transitions to states other than the locally most fit one. This can be seen

as the injection of "noise" or random perturbations into the system, which makes it deviate from its preferred trajectory. Physically, this is usually the effect of outside perturbations (e.g. vibrations, or shaking of the system) or of intrinsic indeterminacy (e.g. thermal or quantum fluctuations, or simply unknown factors that have not been incorporated into the state description). Such perturbations can "push" the system upwards, towards a higher potential. This may be sufficient to let the system escape from a local minimum, after which it will again start to descend towards a possibly deeper valley. For example, if we consider Figure 4, with a system in valley B, a small push to the left may make it cross peak X, and thus enter the basin of the much deeper minimum A.

In general, the deeper the valley, the more difficult it will be for a perturbation to make a system leave that valley. Therefore, noise will in general make the system move out of the more shallow, and into the deeper valleys. Thus, noise will in general increase fitness. The stronger the noise the more the system will be able to escape the relatively shallow valleys, and thus reach a potentially deeper valley. However, a system with noise will never be able to really settle down in a local or global minimum, since whatever level of fitness it reaches it will still be perturbed and pushed into less fit states.

The most effective use of noise to maximize self-organization is to start with large amounts of noise which are then gradually decreased, until the noise disappears completely. The initially large perturbations will allow it to escape all local minima, while the gradual reduction will allow it to settle down in what is hopefully the deepest valley. This is the principle underlying annealing, the hardening of metals by gradually reducing the temperature, thus allowing the metal molecules to settle in the most stable crystalline configuration. The same technique applied to computer models of self-organization is called *simulated annealing*.

6. Conclusion and future prospects

The theory of self-organization and adaptivity has grown out of many disparate scientific fields, including physics, chemistry, biology, cybernetics, computer modelling, and economics. This has led to a quite fragmented approach, with many different concepts, terms and methods, applied to seemingly different types of systems. However, out of these various approaches a core of fundamental concepts and principles has slowly started to emerge which seem applicable to all self-organizing systems, from simple magnets and crystals to brains and societies. The present article has attempted to bring the most important of these ideas together without going into the technical details.

Self-organization is basically the spontaneous creation of a globally coherent pattern out of the local interactions between initially independent components. This collective order is organized in function of its own maintenance, and thus tends to resist perturbations. This robustness is achieved by distributed, redundant control so that damage can be restored by the remaining, undamaged sections.

The basic mechanism underlying self-organization is the deterministic or stochastic variation that governs any dynamic system, exploring different regions in the state space until it happens to reach an attractor, i.e. a configuration that closes in on itself. This process can be accelerated and deepened by increasing variation, for example by adding "noise" to the system. Entering the attractor precludes further variation outside the attractor, and thus restricts the freedom of the system's components to behave independently. This is equivalent to the increase of coherence, or decrease of statistical entropy, that defines self-organization.

Closure sets the system apart from its environment, defining it as autonomous. Closure usually results from the non-linear, feedback nature of interactions. If the feedback is positive it leads to the explosive growth of whatever configuration originally entered the positive feedback regime. This growth ends when all available components have been absorbed into the new configuration, leaving the system in a stable, negative feedback state.

Self-organizing systems have in general several stable states, and this number tends to increase (bifurcate) as an increasing input of energy pushes the system farther from its thermodynamic equilibrium. To adapt to a changing environment, the systems needs a sufficiently large variety of possible stable states to cope with likely perturbations. This variety, however, must not be so large as to make its evolution uncontrollably chaotic. Given this variety, the most adequate configurations are selected according to their fitness, either by the environment, or indirectly by subsystems that have already adapted to the environment at an earlier stage. Thus, the system can adjust its internal configuration to external perturbations, while minimizing the changes to its overall organization.

The theory of self-organization has many potential—but as yet relatively few practical—applications. In principle, it provides an insight in the functioning of most of the complex systems that surround us, from galaxies and planets to molecules, and from living cells to ecosystems and markets. Such an understanding does not necessarily lead to a better capability of prediction, though, since the behavior of self-organizing systems is unpredictable by its very nature. On the other hand, getting a better insight into the relevant sources of variation, selection and intrinsic attractor structures will help us to know which behaviors are likely, and which are impossible.

Managing or controlling self-organizing systems runs into similar limitations: since they intrinsically resist external changes, it is difficult to make them do what you want. Increasing pressure will eventually result in a change, but this may be very different from the desired effect, and may even result in the destruction of the system. The best approach seems to consist in identifying "lever points", that is, properties where a small change may result in a large, predictable effect.

Most practical applications until now have focused on designing and implementing artificial self-organizing systems in order to fulfil particular functions. Such systems have several advantages over more traditional systems: robustness, flexibility, capability to function autonomously while demanding a minimum of supervision, and the spontaneous development of complex adaptations without need for detailed planning. Disadvantages are limited predictability and difficulty of control.

Most such applications have been computer programs, such as neural networks, genetic algorithms, or artificial life simulations, that solve complex problems. The basic method is to define a fitness function that distinguishes better from worse solutions, and then create a system whose components vary relative to each other in such a way as to discover configurations with higher global fitness. For example, the collective behavior of ants, who produce a network of trails that connects their nest in the most efficient way to different food sources, has been a major source of inspiration for programs that try to minimize the load on a heavily used communication network by adapting the routes that the information packets follow to the variable demand.

Although more difficult to control, it is also possible to create self-organizing systems in hardware rather than software. There have been several, partially successful attempts to build autonomous robots or groups of robots functioning according to principles of self-organization. Simpler examples are decorative objects, such as plasma balls that show an

endlessly changing pattern of electrical arborescences, or lamps filled with a mixture of oil and water where a constant heating from below makes oil bubbles move up and down. More critical applications are the creation or restoration of ecosystems. A most spectacular example is Biosphere 2, a completely closed (apart from sunlight) complex ecosystem, built inside a huge greenhouse in the desert of Arizona.

Perhaps the most challenging application would be to design a complex socio-economic system that relies on self-organization rather than centralized planning and control. Although our present organizations and societies incorporate many aspects of self-organization, it is clear that they are far from optimal. Yet, our lack of understanding of social self-organization makes it dangerous to introduce radical changes, however well intended, because of their unforeseen side-effects. Better models and simulations of social systems may be very useful in this respect.

Future developments in the science of self-organization are likely to focus on more complex computer simulations and mathematical methods. However, the basic mechanisms underlying self-organization in nature are still far from clear, and the different approaches need to be better integrated. Although researchers such as Kauffman have started exploring the structure of fitness landscapes for various formally defined systems by computer simulation, we should at the same time try to understand which types of variations, fitness functions and attractor dynamics are most common in natural systems (physical, biological or social), and why. This may help us to focus on those models, out of the infinite number of possible mathematical models, that are most likely to be useful in understanding and managing every-day phenomena.

Acknowledgments

Bibliography

Ashby, W. R., (1952) *Design for a Brain - The Origin of Adaptive Behavior*. Chapman and Hall, London, [an early, but surprisingly modern investigation of mechanisms of adaptation and self-organization]

Bak P. (1996). *How Nature Works: The Science of Self-Organized Criticality*, Springer, Berlin, . [a readable overview of diverse applications of the power laws that govern systems on the edge of chaos].

Haken, H. (1978). *Synergetics: an introduction*, Springer, Berlin. [a review of basic mathematical techniques for modelling self-organization and phase transitions, with examples from physics, chemistry and biology]

Holland J.H. (1996). *Hidden Order: How adaptation builds complexity,* Addison-Wesley . [a short introduction to Holland's work on complex adaptive systems]

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, MIT Press, Cambridge MA. [an in-depth technical investigation of the principles underlying genetic algorithms]

Kauffman S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York, [an extensive, technical review of Kauffman's simulation results with Boolean networks and autocatalytic sets]

Kelly, K. (1994). *Out of Control*, Addison-Wesley, New York [a science journalist details the history, technological applications, and future implications of the science of self-organization and adaptivity].

Nicolis, G, and Prigogine, I. (1977) *Self-Organization in Non-Equilibrium Systems*, Wiley, New York [a general review of the thermodynamic theory of dissipative structures]

Prigogine, I. and Stengers, I. (1984). *Order out of Chaos*, Bantam Books, New York, [a non-technical, philosophical discussion of the implications of self-organization for the scientific world view]

von Foerster, H., (1960). On self-organising systems and their environments, in: *Self-Organising Systems*, M.C. Yovits and S. Cameron (eds.), Pergamon Press, London, pp. 30-50. [original statement of the "order from noise" principle, as part of the proceedings of the first conference on self-organization]

von Foerster, H., and Zopf, G. (eds.) (1962). *Principles of Self-Organization,* Pergamon, New York [proceedings of the second conference on self-organization; includes Ashby's classic paper on the principle of self-organization]

Glossary:

self-organization: the spontaneous emergence of global coherence out of local interactions

adaptivity: a system's capacity to adjust to changes in the environment without endangering its essential organization

statistical entropy: a mathematical measure of the lack of constraint or lack of information about a system's state; equivalent to Shannon's measure of uncertainty

thermodynamic entropy: a measure of the dissipation of energy into heat

(thermodynamic) equilibrium: the static state of minimum energy where no entropy is produced

stationary state: a state characterized by permanent activity

dissipative structure: an organized pattern of activity sustained by the export of entropy out of a far-from-equilibrium system

Bénard rolls: a type of dissipative structure formed by convection between layers in a liquid heated from below

attractor: a region in state space that a system can enter but not leave

basin: the region in state space surrounding an attractor that leads into the attractor

fitness landscape: a representation of the dynamics on a state space where every point has a height corresponding to its fitness value

correlation length: the longest distance over which the components of a system are correlated

bifurcation: the branching of the stable solution to the system of equations that describes a self-organizing system when an order parameter increases

order parameter: a variable that describes the transition between the disordered and ordered regimes

edge of chaos: domain of dynamic activity where complex adaptive systems typically reside, in between the completely ordered, "frozen" regime and the completely disordered, chaotic regime

distributed control: constraint on the overall organization of a system that is not centralized in a distinct subsystem, but held up collectively by all components

boundary conditions: the state of the environment at the boundary of the system insofar as it influences the system's evolution